# Forecasting Research

**Forecasting Research Division**
**Scientific Paper No. 34**

# Atmospheric data assimilation

**by**

**A C Lorenc**

**May 1995**

Meteorological Office
London Road
Bracknell
Berkshire
RG12 2SZ
United Kingdom

Forecasting Research Division

Scientific Paper No 34

# Atmospheric Data Assimilation

by

Andrew C Lorenc[1]

May 1995

© Crown Copyright 1995

Forecasting Research
Meteorological Office
London Road
Bracknell
Berkshire RG12 2SZ
ENGLAND

This paper is based on a lecture given at the Second WMO Symposium on Assimilation of Observations in Meteorology and Oceanography, held in Tokyo, Japan, 13-17 March 1995. It has been submitted for publication in a special issue of the Journal of the Meteorological Society of Japan, dedicated to "Data Assimilation in Meteorology and Oceanography: Theory and Practice".

---

[1] email: aclorenc@meto.govt.uk

# ATMOSPHERIC DATA ASSIMILATION

Andrew C Lorenc
*Meteorological Office, Bracknell, England*

**Abstract**

The assimilation of data into an atmospheric model is described from first principles, with discussion of the various sources of information and modelling assumptions that allow us to supplement otherwise insufficient direct observational data. The process of adding observational information to that already in the model state is discussed from a Bayesian standpoint, with simple examples, leading to a derivation of the standard "OI" and variational analysis equations.

## 1 Introduction

In this paper we discuss the basic physics of the atmospheric data assimilation problem, in order to understand the important factors to be considered in its mathematical solution. The key mathematical technique, the optimal combination of information, is also approached from its Bayesian basics. Much of this is based on earlier papers (Lorenc, 1986, Lorenc and Hammon, 1988). The novelty of this paper is its bringing together of these in a simple didactic form, following the agreed notation of Ide et al. (1995), with very simple examples to aid in the physical interpretation of the analysis equations.

## 2 What is Data Assimilation?

There are insufficient observations at any one time to determine the state of the atmosphere. So if we want a detailed complete picture, we need additional information. This is available as knowledge of the behaviour and probable structure of the atmosphere. For instance the knowledge of the typical structure of a frontal depression enables a human to draw an "analysis" of the atmospheric state, based on scattered observations. To advance beyond this subjective approach, the behaviour of the atmosphere is embodied in a computer model. In particular, knowledge of the evolution with time is embodied in a forecast model. This enables us to use observations distributed in time. The model also provides a consistent means of representing the atmosphere. *Assimilation is the process of finding the model representation which is most consistent with the observations.*

Usually, data assimilation proceeds sequentially in time. The model organises and propagates forward the information from previous observations. The information from new observations is used to modify the model state, to be as consistent as possible with them and the previous information. It is the experience with operational assimilation for NWP that there is usually more information in the model state, from previous observations, than there is in a new batch at a single synoptic time. Thus it is important to preserve this in the assimilation process; it

is not just a question of fitting the new data. Since all information has to be represented within the model, it is important that the model should be of sufficiently high resolution, with physically realistic detail, to represent the information observed.

Some research (see other papers in this volume) is investigating non-sequential data assimilation methods, especially four-dimensional variational assimilation. This paper is designed to provide the groundwork for this, without going into detail.

# 3 Products and Uses of Assimilation

Assimilation produces a convenient, comprehensive, high-resolution, representation of the atmosphere. It has been clearly demonstrated that the use of a computer model is usually better (i.e. leads to better forecasts) than the subjective human approach. The main practical use of these assimilated "analyses" is for initialising numerical weather prediction (NWP) forecasts. They are also useful for climate and general circulation studies, for instance in the calculations of fluxes, which make use of their high resolution and comprehensive coverage. However it must be remembered that the blend of observed and modelled information will vary according to the accuracy and coverage of the observations. So they must be used with great care for model validation, and climate change detection.

Very useful secondary products of a data assimilation system are the statistics on the (mis-)fit of observations to model. These can be more directly used for model (in-)validation, and for the monitoring of observing systems.

# 4 The Optimal Combination of Information

The history of this goes back to Gauss (1809, 1823), who in his study of the motion of the planets developed methods for the weighted combination of observations with errors. We will approach it from a Bayesian standpoint, following Lorenc (1986) and Lorenc and Hammon (1988). The Bayesian formalism gives a consistent treatment allowing for the (unfortunately common) occurrence of observations with "gross" errors. The resulting equations are equivalent to Gauss's minimum variance approach, for observations with a Gaussian error distribution.

## 4.1 Bayes' Theorem for Discrete Events

The Bayesian approach is to use probabilities to describe the accuracy of our knowledge about past events. We then have a formalism for modifying the probabilities in the light of new knowledge; exactly what we need to do in sequential data assimilation. We introduce this with a discrete example for events $A$ $B$. $P(A)$ is the probability of $A$ occurring (this is the usual use of probabilities), or a measure of our certainty that $A$ occurred in the past (this is the Bayesian use of probabilities). Then $P(A \cap B)$ is the probability that $A$ and $B$ both occurred, and $P(A|B)$ is the conditional probability of $A$ given $B$ has occurred. We have two

ways of expressing $P(A \cap B)$:

$$\begin{aligned} P(A \cap B) &= P(B)\ P(A|B) \\ &= P(A)\ P(B|A) \end{aligned} \tag{1}$$

This leads directly to Bayes' Theorem:

$$P(A|B) = \frac{P(B|A)\ P(A)}{P(B)} \tag{2}$$

So we start with a prior probability of $A$, and add the information that $B$ has occurred, to give us the posterior probability of $A$ given $B$. To evaluate this we can calculate $P(B)$ from:

$$P(B) = P(B|A)P(A) + P(B|\bar{A})P(\bar{A}) \tag{3}$$

## 4.2 Single-variable Bayesian Analysis with Gaussian pdfs

For continuous variables we use probability distribution functions (pdfs):

$$p(x)dx = P(\ x \le x^t < x+dx\ ) \tag{4}$$

where $x^t$ is the true value. Bayes' Theorem becomes:

$$p(x|y^o) = \frac{p(y^o|x)p(x)}{p(y^o)} \tag{5}$$

Let us start thinking of $x$ as the model state, which accumulates our knowledge. $p(x)$ is the prior distribution; our knowledge from previous observations. $p(x|y^o)$ is the posterior distribution, after adding the information from the observation $y^o$. $p(y^o|x)$ is the probability density of getting the observation $y^o$, given our previous knowledge. Note that this is a density in $y$-space. Regarded as a function of $x$, $p(y^o|x)$ is not a probability density (its integral is not necessarily one); it is called the likelihood function for $x$.

Let us assume that our prior knowledge is that $x$ is near $x^b$, and that the variance of its deviation from $x^b$ is $V^b$. The usual way of modelling such a distribution is as a Gaussian:

$$p(x) = N(x|x^b, V^b) = (2\pi V^b)^{-\frac{1}{2}} \exp\left(-\frac{1}{2}\frac{(x-x^b)^2}{V^b}\right) \tag{6}$$

If our observation directly measures the variable $x$, with observational error variance $V^o$, then the probability of observed value $y^o$, given the true value is $x$, can also be modelled by a Gaussian pdf:

We can get $p(y^o)$ by integrating over all $x$:

$$p(y^o|x) = N(y^o|x,V^o) = (2\pi V^o)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}\frac{(y-x)^2}{V^o}\right) \tag{7}$$

$$p(y^o) = \int p(y^o|x)p(x)dx \tag{8}$$

For Gaussians (6) and (7) this gives:

$$p(y^o) = N(y^o|x^b,V^o+V^b) \tag{9}$$

Substituting in (5) gives:

$$p(x|y^o) = N(x|x^a,V^a) \tag{10}$$

where

$$\begin{aligned}\frac{x^a}{V^a} &= \frac{y^o}{V^o} + \frac{x^b}{V^b}\\ \frac{1}{V^a} &= \frac{1}{V^o} + \frac{1}{V^b}\end{aligned} \tag{11}$$

This is the standard formula for the combination of observations with error, known since the work of Gauss.

$p(x)$ —the prior distribution, $p(x|y^o)$ —the posterior distribution, and $p(y^o|x)$ —the likelihood function, are plotted in figure 1 for four different values for $y^o$. A unique property of the Gaussian pdfs can be seen in them, and in the above equations: the shape of the posterior distribution, and its relative position to $x^b$ and $y^o$, are both independent of the observed value. They only depend on the error variances.

The reason for this becomes apparent if we take logarithms of both sides of (5). The Bayesian analysis equation becomes:

$$-\ln[p(x|y^o)] = -\ln[p(y^o|x)] - \ln[p(x)] + constant \tag{12}$$

Gaussians become quadratics, which are summed to give another quadratic:

$$\begin{aligned}-\ln[p(x|y^o)] &= \frac{1}{2}\frac{(x^b-x)^2}{V^b} + \frac{1}{2}\frac{(y^o-x)^2}{V^o} + constant.\\ &= \frac{1}{2}\frac{(x^a-x)^2}{V^a} + constant.\end{aligned} \tag{13}$$

The Gaussian curves of figure 1 become the quadratics of figure 2, with the most likely values for $x$ being at the minimum of the total curve. So the Bayesian combination of

Gaussian pdfs gives the same "best" analysis as a weighted least-squares best fit to the data.

## 4.3 One-dimensional Bayesian Analysis

To extend the above to data assimilation of observations into a model, we need to introduce several new ideas. We do this in the context of the simplest relevant example:-

A model which represents the "world" with two grid points:

$$\boldsymbol{x} = \begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \tag{14}$$

An instrument which gives one observed value $y^o$ midway between the points.

We can interpolate a model estimate $y$ of the observed value:

$$\begin{aligned} y &= H(\boldsymbol{x}) = \tfrac{1}{2}x_1 + \tfrac{1}{2}x_2 \\ &= \mathbf{H}\,\boldsymbol{x} = \begin{pmatrix} \frac{1}{2} & \frac{1}{2} \end{pmatrix}\begin{pmatrix} x_1 \\ x_2 \end{pmatrix} \end{aligned} \tag{15}$$

We need a probability distribution (the background pdf) to describe what we know about $\boldsymbol{x}$, before getting the observation. Let us suppose we have a prior estimate $x_1^b$ with error variance $V^b$:

$$p(x_1) = N(x|x_1^b, V^b) = (2\pi V^b)^{-\frac{1}{2}} \exp\left( -\frac{1}{2}\frac{(x_1 - x_1^b)^2}{V^b} \right) \tag{16}$$

Because $\boldsymbol{x}^b$ is produced by a single process, usually a forecast, errors in $x_1^b$ and $x_2^b$ are correlated:

$$\left\langle (x_1^b - x_1^t)(x_2^b - x_2^t) \right\rangle = \mu V^b \tag{17}$$

where $\mu$ is the correlation coefficient. The corresponding joint pdf can be modelled by a multi-dimensional Gaussian:

$$\begin{aligned} p(x_1 \cap x_2) &= p(\boldsymbol{x}) = N(\boldsymbol{x}|\boldsymbol{x}^b, \mathbf{B}) \\ &= ((2\pi)^2 |\mathbf{B}|)^{-\frac{1}{2}} \exp\left( -\frac{1}{2}(\boldsymbol{x} - \boldsymbol{x}^b)^T \mathbf{B}^{-1} (\boldsymbol{x} - \boldsymbol{x}^b) \right) \end{aligned} \tag{18}$$

where $\mathbf{B}$ is the covariance matrix:

$p(\boldsymbol{x})$ is assumed to be the combination of all our prior knowledge about the model state. A

$$\mathbf{B}=V_b\begin{pmatrix}1 & \mu\\ \mu & 1\end{pmatrix}. \tag{19}$$

schematic contour plot, for our simple two parameter model, is shown by the long dashed ellipses in figure 3.

We now consider the new information provided by the observation. A perfect instrument would measure a "true" value $y^t$. Real instruments are not perfect; they have *instrumental error* which we model by a Gaussian. (In our example this is one-dimensional, but we use a more general notation which can be applied if there are more than one observations.)

$$\begin{aligned} p(y^o|y^t) &= N(y^o|y^t,\mathbf{E}) \\ &= (2\pi|\mathbf{E}|)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y^o-y^t)^T\mathbf{E}^{-1}(y^o-y^t)\right) \end{aligned} \tag{20}$$

We cannot use this directly in Bayes' theorem because it is a function of $y^t$, rather than $x$. We need to specify also the probability of a perfect instrument observing $y^t$, given that the true model state is $x$. Because we have to interpolate from $x$ to $y$, we cannot know $y^t$ exactly; the *representativeness error* can be thought of as the error in the interpolation operator **H** in (15).

$$\begin{aligned} p(y^t|x) &= N(y^t|H(x),\mathbf{F}) \\ &= (2\pi|\mathbf{F}|)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y^t-H(x^t))^2\mathbf{F}^{-1}(y-H(x^t))\right) \end{aligned} \tag{21}$$

If we assume that instrumental and representativeness errors are independent, we can convolve them to get a combined *observational error*:

$$\begin{aligned} p(y^o|x) &= \int p(y^o|y^t)p(y^t|x)dy^t \\ &= N(y^o|H(x),\mathbf{E}+\mathbf{F}) \\ &= (2\pi|\mathbf{E}+\mathbf{F}|)^{-\frac{1}{2}}\exp\left(-\frac{1}{2}(y^o-H(x)^T(\mathbf{E}+\mathbf{F})^{-1}(y^o-K(x))\right) \end{aligned} \tag{22}$$

The sum of the instrumental and representativeness error covariances, **E**+**F**, is often written as a single observational error covariance **R**. The above derivation shows that it is composed of two parts, **E** which is a function of the instrument characteristics, and **F** which is a function of the model resolution. For instance, for a wind observation from a radiosonde, the errors in tracking the balloon might lead to an instrumental error of about $1ms^{-1}$. The error of representativeness due to trying to predict such a wind from a model with horizontal grid-length 200km would be about $3ms^{-1}$, and a grid-length on 20km would reduce the error of representativeness to about $1ms^{-1}$.

$p(y^o|x_1,x_2)$ is plotted as a function of $(x_1,x_2)$ in the short dashed lines in figure 3. It is called the likelihood function[2]. Note that the ridge extends to infinity; all values on the ridge line are, as far as the observed information is concerned, equally likely. To get a unique "best estimate", we need to combine this with the prior information.

We can substitute (18) and (22) into the Bayesian analysis equation:

$$\begin{aligned} p(\boldsymbol{x}|y^o) &= \frac{p(y^o|\boldsymbol{x})p(\boldsymbol{x})}{p(y^o)} \\ &\propto N(y^o|H(\boldsymbol{x}),\mathbf{E}+\mathbf{F})\ N(\boldsymbol{x}|\boldsymbol{x}^b,\mathbf{B}) \end{aligned} \tag{23}$$

It is a property of Gaussians that, if $H$ is linearisable :

$$\begin{aligned} &N(y^o|H(\boldsymbol{x}\ ),\mathbf{E}+\mathbf{F})\qquad N(\boldsymbol{x}|\boldsymbol{x}^b,\mathbf{B}) \\ &= N(y^o|H(\boldsymbol{x}^b),\mathbf{E}+\mathbf{F}+\mathbf{H}\mathbf{B}\mathbf{H}^T)\ N(\boldsymbol{x}|\boldsymbol{x}^a,\mathbf{A}) \end{aligned} \tag{24}$$

where $\boldsymbol{x}^a$ and $\mathbf{A}$ are defined by:

$$\begin{aligned} \mathbf{A} &= \mathbf{B}-\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{E}+\mathbf{F})^{-1}\mathbf{H}\mathbf{B} \\ \boldsymbol{x}^a &= \boldsymbol{x}^b+\mathbf{B}\mathbf{H}^T(\mathbf{H}\mathbf{B}\mathbf{H}^T+\mathbf{E}+\mathbf{F})^{-1}(y^o-H(\boldsymbol{x}^b)) \end{aligned} \tag{25}$$

Cancelling the constant of proportionality gives:

$$p(\boldsymbol{x}|y^o) = N(\boldsymbol{x}|\boldsymbol{x}^a,\mathbf{A}) \tag{26}$$

So the posterior probability, after adding the observational information, is a Gaussian with mean $\boldsymbol{x}^a$ and variance $\mathbf{A}$. As in figure 1, the posterior pdf in figure 2 (solid lines) is narrower and taller than the prior. Adding information from the observation increases our confidence, and reduces the error variance from $\mathbf{B}$ to $\mathbf{A}$.

For our simple example the algebra is easily done by hand, giving:

$$\boldsymbol{x}^a = \begin{pmatrix} x_1^a \\ x_2^a \end{pmatrix} = \begin{pmatrix} x_1^b \\ x_2^b \end{pmatrix} + \frac{\left(V^b\left(\frac{1+\mu}{2}\right)\right)^2}{E+F+V^b\left(\frac{1+\mu}{2}\right)}\left[y^o-\frac{x_1^b+x_2^b}{2}\right]\begin{pmatrix}1\\1\end{pmatrix} \tag{27}$$

If this problem is solved using optimal interpolation (OI), the approximation is made that $\mathbf{H}\mathbf{B}\mathbf{H}^T=V^b$, so the $\left(\frac{1+\mu}{2}\right)$ do not appear.

---

[2] It does not integrate to one over $x$, so it is not a probability.

## 4.4 Practical Analysis Methods

We saw in (13), and figure 2, the effect of taking logarithms of the posterior probability. For multi-dimensional Gaussians, this becomes:

$$-\ln[p(x|y^o)] = \frac{1}{2}(x^b-x)^T\mathbf{B}^{-1}(x^b-x) + \frac{1}{2}(y^o-H(x))^T(\mathbf{E}+\mathbf{F})^{-1}(y^o-H(x)) + \textit{constant.} \tag{28}$$

so the most probable $x$ minimises

$$J(x) = \frac{1}{2}(x^b-x)^T\mathbf{B}^{-1}(x^b-x) + \frac{1}{2}(y^o-H(x))^T(\mathbf{E}+\mathbf{F})^{-1}(y^o-H(x)) \tag{29}$$

For linearisable $H$ we can solve this explicitly, giving (25). This is known as the "OI" (Optimal Interpolation) analysis equation; but since it is only optimal if the assumptions about the pdfs are correct, a preferred name is Statistical Interpolation. The same equations appears in the analysis step of the Extended Kalman Filter (Ghil and Malanotte-Rizzoli, 1991).

(25) requires the solution of a linear system of order the number of observational data. Meteorological observations are so numerous that practical simplifications are required; for instance by partitioning into analysis volumes, and modelling $\mathbf{HBH}^T$ by a correlation function (Lorenc 1981). Alternatively, iterative methods can be used. One such, the "Successive Correction Method", was developed pragmatically, before OI. It can be derived from (25) if we define:

$$\mathbf{W} = \mathbf{BH}^T(\mathbf{E}+\mathbf{F})^{-1} \tag{30}$$

$$\mathbf{Q} \approx (\mathbf{I}+\mathbf{WH})^{-1} \tag{31}$$

$$x^{[u+1]} = x^{[u]} + \mathbf{Q}(\mathbf{W}(y^o-H(x^{[u]})) + x^b-x^{[u]}) \tag{32}$$

Lorenc (1992) showed that, particularly for cases where the correlation function used to model **B** has a simple form, an easily calculated approximation to **Q** is sufficient to get iteration (32) to converge to the "OI" result.

Variational Analysis (Var) methods search directly for the minimum of (29), using a descent algorithm which needs the gradient of $J$ :
If (33) is calculated accurately, they can use more sophisticated and efficient descent

$$\left(\frac{\partial J}{\partial x}\right)^T = -\mathbf{B}^{-1}(x^b - x) - \mathbf{H}^T(\mathbf{E}+\mathbf{F})^{-1}(y^o - H(x)) \tag{33}$$

algorithms than the pre-conditioned steepest descent which is used in (32). The prediction ($H$) of observed values y from a model state x, can be generalised from a simple interpolation, as in (15), to a more general prediction using a forecast model and observation operators, allowing Var to be applied to observations distributed in time, and indirect observations (e.g. satellite radiances).

# 5 Non-Gaussian PDFs

## 5.1 Model for Observational Errors

The simplest model that allows for the observed fact that observational errors are not in practice Gaussian, is to assume that a small fraction of the observations are corrupted, and hence worthless. The others have Gaussian errors. For each observation we have:

$$p(y^o|x) = p(y^o|G\cap x)P(G) + p(y^o|\bar{G}\cap x)P(\bar{G}) \tag{34}$$

$G$ is the event "there is a gross error".[3]
where $\bar{G}$ means *not* $G$.

$$p(y^o|\bar{G}\cap x) = N(y^o|H(x),\mathbf{E}+\mathbf{F})$$

$$p(y^o|G\cap x) = \begin{cases} k & \textit{over the range of plausible values} \\ 0 & \textit{elsewhere} \end{cases} \tag{35}$$

## 5.2 Analysis, allowing for gross errors

The results of assuming such a pdf can be quite dramatic, even if $P(G)$ is small. Figure 4 shows the equivalent of figure 1, with errors appropriate for pressure observations from ships, which have about 5% gross errors. When the observation and the background agree, there is little difference from figure 1. But when they disagree, the posterior distribution becomes bi-modal. It is worth noting that, in contrast to the statement at the end of section 4.3, the posterior variance can be greater than the prior variance - a strange result for those used to dealing with Gaussian distributions.

The normal way of dealing with non-Gaussian, long-tailed distributions such as these is to detect and reject probable gross errors in a quality control step prior to the analysis, and then to assume that the errors of the remaining observations come from a Gaussian distribution. The quality control decision can either use pragmatic criteria, or apply the discrete Bayes'

[3] We assume that $G$ is independent of $x$, so that $P(G|x)=P(G)$.

theorem to the event $G$ "there is a gross error" (Lorenc and Hammon, 1988):

$$P(G|y^o) = \frac{P(y^o|G)\ P(G)}{P(y^o)} \tag{36}$$

Figure 5 shows the figure 4 examples in the log(probability) form of figure 2. It illustrates the possibility, and problems, of using a non-Gaussian distribution directly in a variational scheme. The observational penalty is not quadratic; it has plateaus away from the observed value. Adding this to a quadratic background penalty can give multiple minima, making it uncertain that a descent algorithm will find the absolute minimum.

# 6 Useful prior knowledge about the atmosphere

Having discussed the mathematical problems of the optimal combination of information, and derived some analysis equations, let us return to discuss the physical problem of the sorts of prior knowledge about the atmospheric behaviour which might supplement the observational information.

We need first to decide what aspects of the atmosphere we are modelling. Certain scales, including of course all those not resolved by the model, and type of motion, e.g. sound waves, are not represented. We define our "true" state, the target of an ideal assimilation, as the atmospheric state with these filtered out. They then contribute to the error of representativeness of the model, but not to the analysis error (i.e. if the analysis fits our "true" state, its analysis error is zero, even if it does not represent all the detail of the real atmosphere).

For the scales considered in NWP, the atmosphere is usually smooth, slowly varying, and close to horizontal non-divergence. We can observe these properties and quantify them. The smoothness can be described by a correlation function, or power spectrum. The slowly varying property leads to useful balance relationships. For instance by assuming that the rate of change of wind, and other residual terms, are negligible compared to the pressure gradient and coriolis forces, we get the geostrophic relationship. Unfortunately none of these simple relationships is exact, and imposing them would probably damage our prior estimate of the atmospheric state, $\boldsymbol{x}^b$. To avoid this, we apply the same balance arguments to $\boldsymbol{x}^b$, and subtract. The residual terms which lead to inaccuracy in the simple relationships then approximately cancel. It is thus appropriate to apply the smoothness, geostrophic balance, and non-divergence relationships to the residuals, $\boldsymbol{x}$-$\boldsymbol{x}^b$. They are used in modelling the structure of the background covariance $\mathbf{B}$.

The observation that the atmosphere (for those aspects we are trying to represent) is slowly varying, can be used to derive balance relationships involving diabatic processes, such as latent heating, and the circulations of the associated weather systems. However these effects are not easily described mathematically. An alternative approach is to use directly the

"slowly varying" property by *initialising* a numerical model so that its initial evolution is slowly varying.

By far the most useful and accurate prior knowledge that we have, is of the equations governing the atmosphere's evolution in time, enabling us to build NWP models. The simplest use of models is to carry information forward in time from a past analysis, to provide the background for a new analysis. This process is the core of all operational NWP systems; without it analyses and forecasts would be substantially less accurate. Computationally feasible methods for integrating this knowledge more effectively into the assimilation, treating it as a single four-dimensional problem, are the subject of current research.

## References


Ghil, Michael, and Paola Malanotte-Rizzoli, 1991
"Data Assimilation in Meteorology and Oceanography". *Advances in Geophysics*, **33**, 141-266

Lorenc, A.C. 1981
"A global three-dimensional multivariate statistical analysis scheme." *Mon. Wea. Rev.*, **109**, 701-721.

Lorenc, A.C. 1986
"Analysis methods for numerical weather prediction." *Quart. J. Roy. Met. Soc.*, **112**, 1177-1194.

Lorenc, A.C. and Hammon, O., 1988
"Objective quality control of observations using Bayesian methods - Theory, and a practical implementation." *Quart. J. Roy. Met. Soc.*, **114**, 515-543

Lorenc, Andrew 1992
"Iterative analysis using covariance functions and filters" *Quart. J. Roy. Met. Soc.*, **118**, 569-591

## Legends for figures

Figure 1. Prior pdf $p(x)$ *(dashed line)*, posterior pdf $p(x|y^o)$ *(solid line)*, and likelihood of observation $p(y^o|x)$ *(dotted line)*, plotted against $x$ for various values of $y^o$. (Adapted from Lorenc and Hammon 1988).

Figure 2. As figure 1 for -log(probabilities).

Figure 3. Contour plots for the simple example of section 5.2:
long dashes: the prior pdf $p(x_1,x_2)$, a two-dimensional Gaussian with centre $(x_1^b,x_2^b)$,
short dashes: the likelihood function $p(y^o|x_1,x_2)$, a Gaussian ridge about the line $\frac{1}{2}(x_1+x_2)=y^o$,
thick lines: the posterior pdf $p(x_1,x_2|y^o)$, a two-dimensional Gaussian with centre $(x_1^a,x_2^a)$.

Figure 4. As figure 1 for an observation with a 5% chance of gross error.

Figure 5. As figure 4 for -log(probabilities).

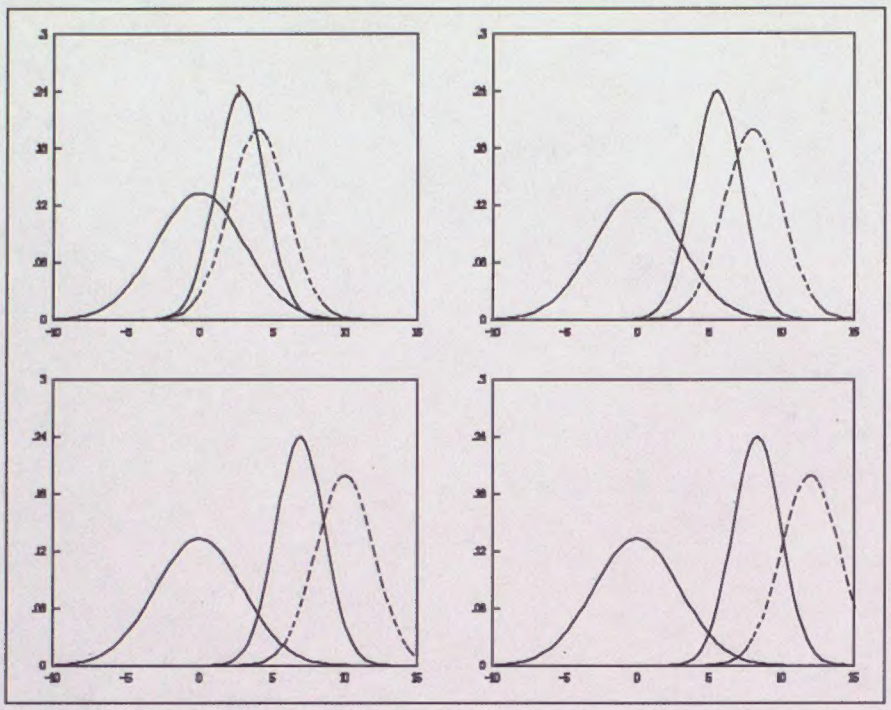

Figure 1. Prior pdf $p(x)$ *(dashed line)*, posterior pdf $p(x|y^o)$ *(solid line)*, and likelihood of observation $p(y^o|x)$ *(dotted line)*, plotted against $x$ for various values of $y^o$. (Adapted from Lorenc and Hammon 1988).

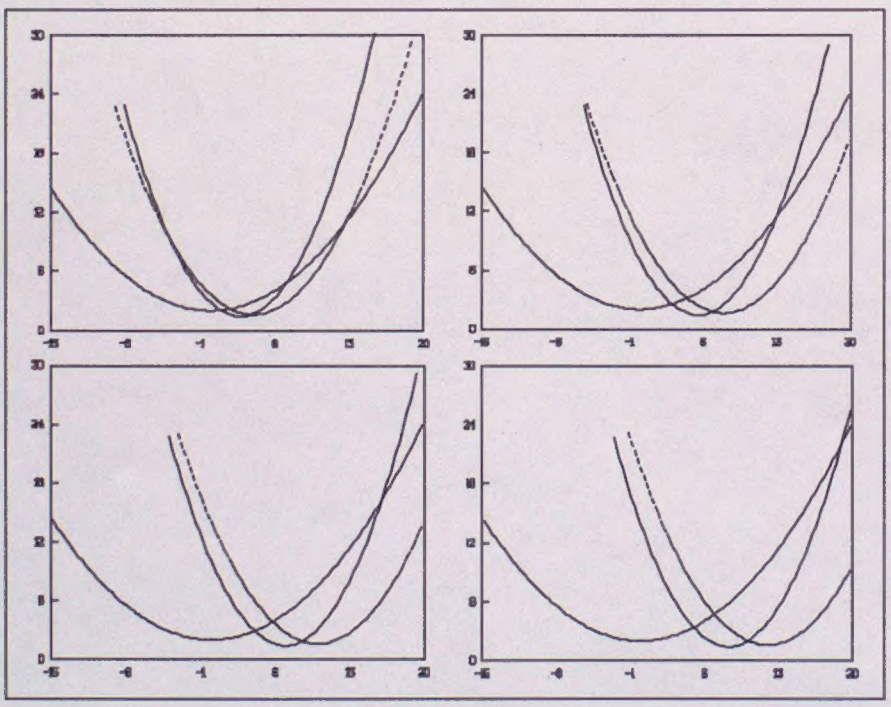

Figure 2. As figure 1 for -log(probabilities).

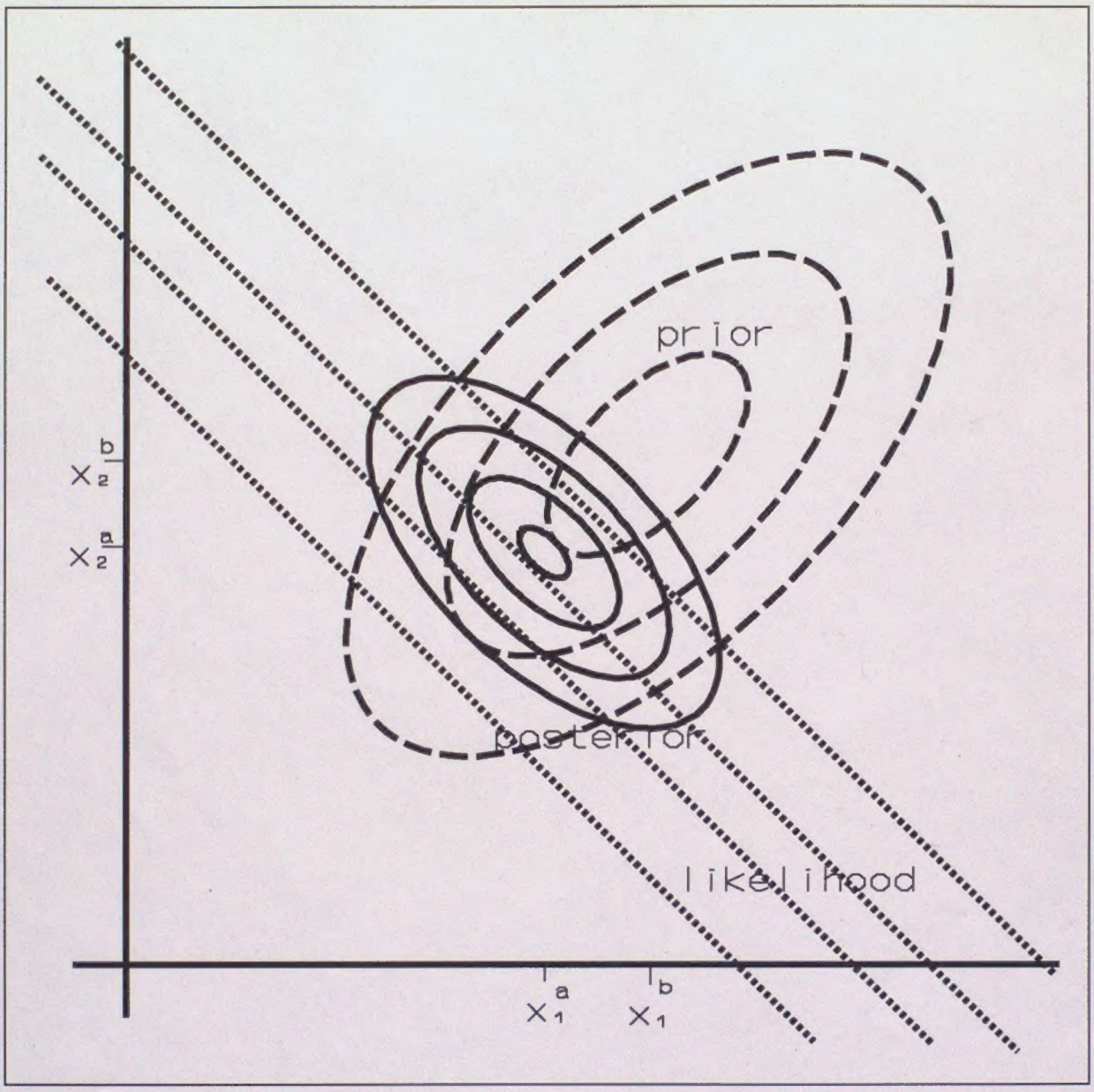


Figure 3. Contour plots for the simple example of section 5.2:
long dashes: the prior pdf $p(x_1,x_2)$, a two-dimensional Gaussian with centre $(x_1^b,x_2^b)$,
short dashes: the likelihood function $p(y^o|x_1,x_2)$, a Gaussian ridge about the line $\frac{1}{2}(x_1+x_2)=y^o$,
thick lines: the posterior pdf $p(x_1,x_2|y^o)$, a two-dimensional Gaussian with centre $(x_1^a,x_2^a)$.

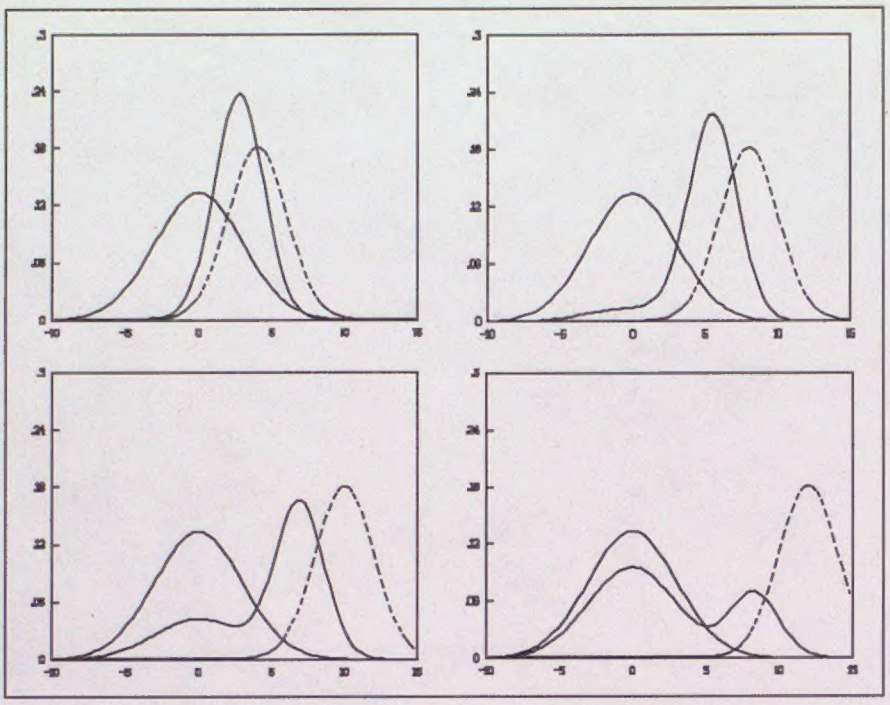

Figure 4. As figure 1 for an observation with a 5% chance of gross error.

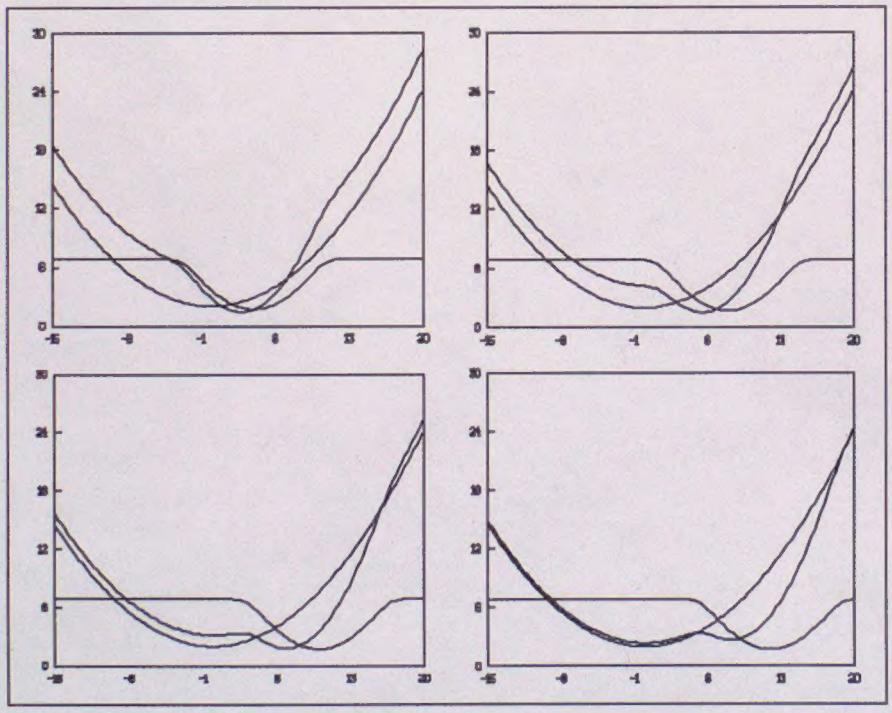

Figure 5. As figure 4 for -log(probabilities).